

(2) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2004 年 2 月 12 日 (12.02.2004)

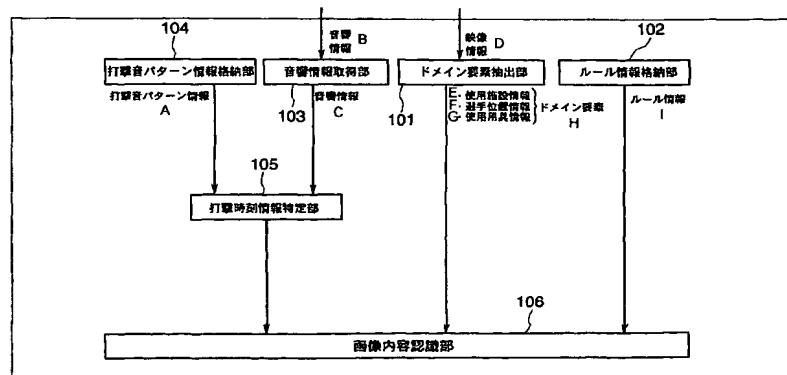
PCT

(10) 国際公開番号
WO 2004/013812 A1

- (51) 国際特許分類⁷: G06T 7/20, 7/60, 1/00 金井市 貫井北町 4-2-1 独立行政法人通信総合研究所内 Tokyo (JP).
- (21) 国際出願番号: PCT/JP2002/007826
- (22) 国際出願日: 2002 年 7 月 31 日 (31.07.2002)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人 (米国を除く全ての指定国について): 独立行政法人通信総合研究所 (COMMUNICATIONS RESEARCH LABORATORY, INDEPENDENT ADMINISTRATIVE INSTITUTION) [JP/JP]; 〒184-8795 東京都小金井市貫井北町 4-2-1 Tokyo (JP).
- (74) 代理人: 赤澤 一博 (AKAZAWA, Kazuhiro); 〒604-8161 京都府京都市中京区烏丸通六角上ル饅頭屋町617 六角ビル6F Kyoto (JP).
- (81) 指定国 (国内): AU, CN, JP, KR, US.
- (84) 指定国 (広域): ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR).
- 添付公開書類:
— 国際調査報告書
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 宮森 恒 (MIYAMORI, Hisashi) [JP/JP]; 〒184-8795 東京都小
- 2 文字コード及び他の略語については、定期発行される各 PCT ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(54) Title: IMAGE RECOGNITION APPARATUS AND IMAGE RECOGNITION PROGRAM

(54) 発明の名称: 画像認識装置及び画像認識プログラム



104... STRIKE SOUND PATTERN INFORMATION STORAGE BLOCK
 A... STRIKE SOUND PATTERN INFORMATION
 B... ACOUSTIC INFORMATION
 C... ACOUSTIC INFORMATION
 D... VIDEO INFORMATION
 101... DOMAIN ELEMENT EXTRACTION BLOCK
 E... FACILITIES INFORMATION
 F... PLAYER POSITION INFORMATION
 G... INSTRUMENT INFORMATION
 H... DOMAIN INFORMATION
 102... RULE INFORMATION STORAGE BLOCK
 I... RULE INFORMATION
 106... IMAGE CONTENT RECOGNITION BLOCK
 103... ACOUSTIC INFORMATION ACQUISITION BLOCK
 105... STRIKE TIME INFORMATION IDENTIFICATION BLOCK

(57) Abstract: An image processing apparatus for recognizing behavior of a player from a content which has recorded a sport played between regions partitioned by an obstacle such as a net. The apparatus includes a video information extraction block for extracting video information having at least video of one of the players, an acoustic information extraction block (103) for extracting from the content, acoustic

[続葉有]



information synchronized with the video information such as a strike sound generated when an instrument such as a ball moving between the regions is struck, a strike time information identification block (105) for identifying the strike time when the instrument is struck according to the acoustic information, a rule information storage block (102) for storing rule information of the sport, and an image content recognition block (106) for recognizing the image content including the player behavior indicated by the video information according to the video information, the position of the instrument at the strike time, and the rule information.

(57) 要約:

ネット等の障害物で区画された領域間で対戦するスポーツを記録したコンテンツから選手の動作を認識する画像処理装置であって、前記コンテンツから少なくとも一方の選手の動作が映った映像情報を取得する映像情報取得部と、前記領域間を移動するボール等の使用用具の打撃時に発生する打撃音等の前記映像情報と同期した音響情報を前記コンテンツから取得する音響情報取得部 103 と、前記音響情報に基づき前記使用用具を打撃した打撃時刻を特定する打撃時刻情報特定部 105 と、当該スポーツのルール情報を格納するルール情報格納部 102 と、前記映像情報と前記打撃時刻における使用用具の位置と前記ルール情報とに基づき、その映像情報が示す選手の動作を含む画像内容を認識する画像内容認識部 106 を備える。

明細書

画像認識装置及び画像認識プログラム

5 技術分野

本発明は、放映されたスポーツ番組等スポーツに関するコンテンツにおいて、従来認識が困難であった画像内容に対しても好適に画像認識できる画像認識装置に関するものである。

10

背景技術

近年、インターネット社会の発展とともに、コンピュータ機器、通信環境、インタフェースが高速化、広帯域化し、様々な映像データが大量に随所に蓄積されるようになる等、さまざまな分野においてユーザが利用できる
15 デジタル映像情報が増加の一途をたどっており、これら膨大な情報にアクセスし映像中の自分の見たい個所を素早く探し出す技術がますます重要視されている。

例えば、テニスなどのスポーツ映像の各場面から、利用者が希望する画面を抽出する場合、「パッシング成功」や「スマッシュ成功」といった画像内容を認識するための方法として、例えば、手入力によって映像情報のどの区間が「パッシング成功」か「スマッシュ成功」かを
20 逐次入力して画像内容を認識したり、コンピュータによってボール、選手、コートラインのそれぞれの位置を抽
25

出しそれら空間的相対関係の時間変化を総合的に判定することにより画像内容を認識したりする方法が考えられる。

ところが、手入力による画像認識を行う場合には、
5 実に画像内容を認識することができるが、人件費が大きくなったり長時間のコンテンツを処理する場合には作業者に大きな負担を与えてしまったりするといった問題点があった。また、コンピュータによる自動認識で画像認識を行う場合には、映像情報のみを処理対象とすると、
10 選手やネット等にボールが重なったり隠れたりした際にそのボールの追跡を失敗し、重要な位置や時刻を特定できない部分が生じ、結果的に画像認識すべきイベントを検出できなかったり誤った画像認識を行うといった不具合があった。

15

発明の開示

上記の課題を解決するために、本発明は、次のような手段を講じたものである。

すなわち、本発明は、ネット等の障害物で区画された
20 領域間で対戦するスポーツにおける選手の動作を、そのスポーツの放映中の番組又は放映前の素材映像やVTR等の記録媒体に記録されるコンテンツから認識する画像認識装置であって、前記コンテンツから少なくとも一方の選手のプレイ中の動作が映った映像情報を取得する映
25 像情報取得部と、前記領域間を移動し当該スポーツの得

点のカウント対象となるボール等の使用用具の打撃時に発生する打撃音等前記映像情報と同期した音響情報を前記コンテンツから取得する音響情報取得部と、この音響情報取得部が取得した音響情報に基づき前記使用用具を打撃した打撃時刻を特定する打撃時刻情報特定部と、当該スポーツを行うためルール情報を格納するルール情報格納部と、前記映像情報取得部で取得した映像情報と前記打撃時刻情報特定部で特定した打撃時刻における使用用具の位置と前記ルール情報格納部に格納されるルール情報とに基づき、その映像情報が示す選手の動作を含む画像内容を認識する画像内容認識部とを備えていることを特徴とする。

このようなものであれば、映像中において使用用具が選手やネット等の障害物に重なったり隠蔽されたりして使用用具の位置特定が困難なとき等、映像情報だけでは画像認識が困難な場合でも、音響情報取得部が取得した打撃音を含む音響情報に基づき打撃時刻情報特定部がその打撃音の発生した打撃時刻を特定し、さらに、特定した打撃時刻と選手のプレイ中の動作が映った映像情報と当該ルールを行うためのルール情報とに基づき画像内容認識部が確実に選手の動作を特定するため、例えば、重なりや隠蔽によるフォアハンドスイング、バックハンドスイング及びオーバーヘッドスイングの認識誤りが生じることの無い画像認識に優れた画像認識装置を提供することができる。

なお、打撃時刻を特定する方法としては、前記打撃時刻情報特定部が、前記音響情報が、所定のレベルより大きな値を示した際に、この大きな値を示す時刻を打撃時刻と特定する方法が挙げられる。

- 5 また、音響情報に含まれる打撃音以外のノイズ音を除去するためには、前記音響情報取得部が、所定の周波数帯域を通過させるフィルタ部を設け、前記音響情報がこのフィルタ部を通過したものであることが望ましく、特に、プレイ中の選手のシューズとコートとが擦れた際に
10 生じる音や風の音やその他雑音等の環境音を好適に除去するためには、前記フィルタ部が、バンドパスフィルタで構成されることが望ましい。

- さらに、効率よく打撃時刻を特定するためには、前記音響情報から抽出した前記打撃音を含む所定時間を有する
15 打撃音候補データに基づき、前記打撃時刻情報特定部が打撃時刻を特定することが好ましい。

- また、確実に打撃時刻を抽出するためには、一の時刻の打撃音候補データとその次の時刻の打撃音候補データとが相互に重なる時刻を有するように前記音響情報から
20 複数個の打撃音候補データを抽出し、この複数個の打撃音候補データに基づき、前記打撃時刻情報特定部が打撃時刻を特定すればよい。さらにこのとき、前記複数個の打撃音候補データを各々同一のデータ長を有するように構成すると共に、複数個の打撃音候補データを前記音響
25 情報から一定時間間隔で抽出するように構成すれば、打

撃音を効率的に抽出できる。

さらに、打撃音の発生した時刻をより確実に確定するためには、前記使用用具と選手がプレイ中常に持って使用するラケット等の用具との当たり具合等による音の変化をパターン化した打撃音パターン情報を格納する打撃音パターン情報格納部を備え、前記打撃時刻情報特定部が、この打撃音パターン情報格納部に格納される打撃音パターン情報と前記音響情報とに基づき前記打撃時刻を特定することが望まれる。

10 また、コンテンツから選手の特徴的な動作を抽出するためには、前記映像情報取得部が、ネット等の障害物や前記領域及びその領域外の境界を示す境界線等の使用施設情報と選手の位置を示す選手位置情報と前記領域間を移動し当該スポーツの得点のカウント対象となる使用用具情報とを映像情報から抽出するドメイン要素抽出部を
15 備えていることが好ましい。

この場合、コンテンツから選手の特徴的な動作をより効率的に抽出するためには、前記選手位置情報が、選手とその選手がプレイ中常に持って使用する用具とを含む
20 領域を示す位置情報であることが望まれる。

さらに、本発明において前記映像情報から選手位置情報を抽出する具体的な態様としては、前記ドメイン要素抽出部が、該ドメイン要素抽出部で抽出した使用施設情報に基づき前記映像情報から選手位置情報を抽出する方
25 法が挙げられる。また、前記映像情報から使用用具情報

を抽出する具体的な態様としては、前記ドメイン要素抽出部が、該ドメイン要素抽出部で抽出した使用施設情報及び選手位置情報に基づき前記映像情報から使用用具情報を抽出する方法が挙げられる。

- 5 また、コンテンツからその分野に関するコンテンツ要素を好適に抽出するためには、前記使用施設情報、前記選手位置情報、前記使用用具情報及び前記ルール情報が、画像抽出の対象となるスポーツ種目に関する知識に基づくものであればよい。

10

図面の簡単な説明

第1図は、本発明の実施形態における画像認識装置の機器構成図である。

第2図は、同実施形態における機能ブロック図である。

- 15 第3図は、同実施形態における映像情報からコートラインの抽出に用いるコートモデルを示す図である。

第4図は、同実施形態における映像情報からネットラインの抽出に用いるネットモデルを示す図である。

- 20 第5図は、同実施形態における映像情報から抽出したコートライン及びネットラインを示す図である。

第6図は、同実施形態における選手領域の検出を示す図である。

第7図は、同実施形態におけるボール領域の検出を示す図である。

- 25 第8図は、同実施形態におけるボール位置の追跡を示す

図である。

第 9 図は、同実施形態におけるルール情報格納部の格納態様を示す図である。

第 10 図は、同実施形態における選手の動作を識別する態様を示す図である。

第 11 図は、同実施形態における映像情報から画像認識を行う過程を示すフロー図である。

発明を実施するための最良の形態

10

以下、本発明の一実施形態について図面を参照して説明する。

第 1 図は、本実施形態における画像認識装置の機器構成を示す機器構成図である。第 2 図は、同実施形態における機能ブロック図である。

本実施形態に係る画像認識装置は、テレビジョン受像機 TV や VTR 等の記録再生装置を用いて表示する放映中の番組や記録媒体などに記録されるスポーツに関するコンテンツから、特徴的な試合中の選手の動作を認識するものであって、第 1 図に示すように、テレビジョン受像機 TV や VTR 等の記録再生装置と接続される入出力インタフェース 11 と、データやプログラム等を記憶する HDD 等の外部記憶装置 12 や内部メモリ 13 と、前記外部記憶装置 12 等に記憶させたプログラムにしたがって作動し当該装置を画像認識装置 1 として機能させる

C P U 1 4 と、ユーザの利用者情報を受け付けるキーボードやマウス等のユーザインタフェイス 1 5 等とを主な構成要素としている。なお、ここで「コンテンツ」とは、選手の動作、コート斜め上から縦方向にコートが映る
5 ようなアングルで撮影されたショット、審判、観客のアップとなるショット等を含む映像や解説者等の音声を含むものとしている。本実施形態では、「コンテンツ」の一例としてテニス番組を例に挙げて説明する。

一方、この画像認識装置 1 を機能面で説明すると、前
10 記 C P U 1 4 等が動作することにより、第 2 図に示すように、ドメイン要素抽出部 1 0 1、ルール情報格納部 1 0 2、音響情報取得部 1 0 3、打撃音パターン情報格納部 1 0 4、打撃時刻情報特定部 1 0 5、画像内容認識部 1 0 6 等としての機能を有している。

15 以下、これら各部を詳述する。

ドメイン要素抽出部 1 0 1 は、ネット等の障害物や区画された領域たるコート及びそのコート外の境界を示す境界線たるコートライン等の使用施設情報と、選手の位置を示す選手位置情報と、前記コート間等を移動し当該
20 スポーツの得点のカウント対象となる使用用具情報とをテレビジョン受像機に表示される映像情報から抽出するものであって、前記コンテンツから少なくとも一方の選手のプレイ中の動作が映った映像情報を取得する映像情報取得部としての機能の一部を発揮するように構成され
25 ている。なお、本実施形態では、抽出する使用施設情報

をコートラインとネットラインとし、また、抽出する選手位置情報を対戦するプレイヤー1の位置情報とプレイヤー2の位置情報とし、さらに、抽出する使用用具情報をテニスボール（以下、「ボール」と称する。）としている。

- 5 さらに、このドメイン要素抽出部101で抽出した使用施設情報、選手位置情報、使用用具情報を、以下、ドメイン要素と総称する。

より具体的には、使用施設情報の抽出は、第3図に示すように、コートラインの代表的な点を示すコート特徴
10 点 P_{c_1} 、 \dots 、 $P_{c_{14}}$ （以下、「 P_c 」と総称する。）
、コートライン L_{c_1} 、 \dots 、 L_{c_9} （以下、「 L_c 」
と総称する。）を定めたコートモデルと、第4図に示すように、ネットラインの代表的な点を示すネット特徴点 P_{n_1} 、 \dots 、 P_{n_3} （以下、「 P_n 」と総称する。）
15 トライン L_{n_1} 、 L_{n_2} （以下、「 L_n 」と総称する。）を
定めたネットモデルとを参照し、コートライン、ネット
ラインの順に映像情報から抽出するように設定している
。

まず、コートラインの抽出は、映像情報から前記コ
20 ト特徴点を検出することによって抽出するようにしてい
る。さらに詳述すると、時刻 $t = 0$ において、初期特徴
点 $P_c(0)$ を入力として与え、次に、 $P_c(0)$ で決
まるコートライン $L_c(0)$ について、各ラインをハフ
平面に変換し、ハフ平面上における各ピーク点を中心に
25 大きさ w_{th} 、 w_r の検出窓 $W_c(0)$ を用意するよう

に設定している。また、時刻 $t = t$ において、まず、原画像の2値化画像 $B(t)$ とコートライン $L_c(t-1)$ の近傍領域とのANDをとり、コート近傍のみからなる2値化画像（以下、「コートライン2値画像」と呼ぶ。

5) $B_c(t)$ を生成するように設定している。そして、これを各ライン毎にハフ変換し、各検出窓 $W_c(t-1)$ で制限される範囲でピーク検出を行い、コート特徴点 $P_c(t)$ を更新し、再び、コートライン $L_c(t)$ をハフ変換し、検出窓 $W_c(t)$ も更新することにより、

10 映像情報からコートラインを抽出するように設定している。なお、パンニングなどによってコート特徴点が画面から外れた場合には、コート中央のコート特徴点 $P_{c_i}(t)$ ($i = 9, 10, 12, 13$ or $10, 11, 13, 14$) は画面内に常に映っていることを仮定し、連結

15 知識を用いて画面外の点を推定して更新するように設定している。同様の理由で、いくつかの初期特徴点は省略してもよい。なお、連結知識とは、例えば、コート中央のコート特徴点 $P_{c_i}(t)$ ($i = 9, 10, 12, 13$) を連結すればコートモデル上に意味付け可能な区画を

20 構成できるといった当該スポーツを行う知識に基づき定義されるものである。

次に、ネットラインの抽出は、時刻 $t = 0$ での初期特徴点 $P_n(0)$ を入力として与え、コートラインと同様に、各ライン毎にネットライン $L_n(0)$ 、検出窓 $W_n($

25 $0)$ を用意し、一方、時刻 $t = t$ においては、原画像の

2 値化画像からコートライン 2 値画像を除いた画像 $B_n(t) = B(t) - B_c(t)$ をネットライン 2 値画像として生成し、これを用いてハフ変換、検出窓内でピーク検出を行って特徴点 $P_n(t)$ を更新し、映像情報からネットラインを抽出するように設定している。

以上のようにして、第 5 図に示すように、コートライン及びネットラインを抽出することができる。

次に、映像情報からこれらコートライン及びネットラインを除去した 2 値画像において重なりが最大となる領域を特定することで選手位置情報を抽出するように構成している。

より具体的には、時刻 $t = t$ において前後 s フレーム離れた画像との差分を求め適当な閾値で 2 値画像 $B_1(t)$ 、 $B_2(t)$ を生成する。なお、ここで、 $B_1(t) = \text{BIN}(I(t) - I(t - s))$ 、 $B_2(t) = \text{BIN}(I(t + s) - I(t))$ としている。ここで BIN は () 内の引数を 2 値化することを示す関数である。そして、これら 2 つの差分画像の AND 演算を行った結果得られる 2 値画像を $B_{diff}(t)$ と、時刻 $t = t$ における画像 $I(t)$ の点で予め用意した選手のユニフォーム等の代表色に対応する色グラスタ内に含まれるものを 1 とした 2 値画像 $B_{label}(t)$ とに基づきコートライン及びネットラインを消去する。さらに、選手領域との重なり部分を除去したと考えられる領域については拡大縮小処理によって領域を補完する。そしてこのように

して得られた 2 つの画像の OR 演算を行い、第 6 図に示すような 2 値画像 $B(t)$ を得る。このようにして得た 2 値画像 $B(t)$ 内で連結領域のラベリングを行い、ノイズの影響を避けるため、これを数フレーム観測し、コート近傍内にある一定以上の面積をもつ領域を選手初期位置とする。そして、時刻 $t = t$ における一定以上の面積をもつ領域のうち、時刻 $t = t - 1$ の選手領域の近傍に存在し、かつ、面積差が最も小さい領域を時刻 $t = t$ における選手領域 p と判定し、選手位置情報を得られるように設定している。

そして、このようにして抽出した選手位置情報との距離に応じて、検出モードと追跡モードとを切り替えることでボールを抽出するように構成している。

さらに詳述すると、検出モードとは、予め用意した大きさ $b_x \times b_y$ のボールを含む所定のテンプレート $T_b(x, y)$ に基づき、第 7 図に示すように、時刻 t において選手領域 p を消去した画像 I'_B 内の選手近傍領域において、テンプレート $T_b(x, y)$ に一致するボール候補 Ba 位置を全て検出するものであって、時刻 $t = t + 1, t + 2, \dots$ についても同様にボール候補を検出し、選手位置を中心に放射状に連続して検出されるボール候補 Ba を絞り込み、最終的に残った 1 つのボール候補 Ba の組みをその時区間のボール軌道 BW と特定することができる。ここで、前記テンプレート $T_b(x, y)$ とは、映像情報からボールを抽出するために設ける一

種のツールであって、本実施形態では、映像中に拡大あるいは縮小表示されるボールの大きさを暫定的に $b_x \times b_y$ とし、この $b_x \times b_y$ から外周を外側へ若干拡大したものをテンプレートとして設定している。

- 5 追跡モードとは、前記テンプレート $T_b(x, y)$ とのテンプレートマッチングによりボール軌道 BW を追跡するものである。ただし、ボール軌道 BW は微小時間でほぼ直線とみなせると仮定し、前回検出された移動量をそのまま現フレームに加えた位置を予測中心として探索
- 10 を行う。そして、選手領域とボール候補 Ba 位置との距離がある閾値より小さくなったら、前記検出モードを実行し、そうでなければ、この追跡モードを繰り返し行うように設定している。

- そして、以上のようにして、第8図に示すように、任意の時間区間におけるボール軌道 BW を得ることができる。なお、第8図は、ボールの軌道 BW をわかりやすく表現するために、任意の時刻の映像情報に重ねて表示している。
- 15

- ルール情報格納部 102 は、当該スポーツを行うために必要なルール情報を格納するものであって、前記外部記憶装置 12 や内部メモリ 13 の所定領域に形成している。より具体的にこのルール情報は、例えば、第9図に示すように、インデックス化したルール情報インデックス「サービス」をルール情報「サービスを開始する直前
- 20
- 25 、サーバーはベースラインのネットに向かって後方、セ

ンターマークとサイドラインの仮想延長線間に両足とも地面につけて立つ。手でボールを空中いずれの方向にでもほうり投げ、そのボールが地面に落ちるまえにラケットで打つ。サービスはラケットとボールが接した瞬間に
5 完了したものとみなす。」と定義し、ルール情報インデックス「コートライン上の落球」をルール情報「コートライン上に落ちたボールは、そのコートラインによって区切られたコート内に落ちたとみなされる。」と定義したものの等を格納している。

10 音響情報取得部 103 は、ボールの打撃時に発生する打撃音等を含む音響情報を前記コンテンツから取得するものであって、この音響情報を、分解能 16 bit、サンプリングレート 44.1 kHz にてサンプリングして取得するように設定している。また、本実施形態では、
15 この音響情報取得部 103 に図示しないフィルタ部を設け、例えば、プレイ中の選手のシューズとコートとが擦れた際に生じる音や風の音やその他雑音等打撃音以外の音響情報をフィルタリングして打撃音のみを好適に抽出できるように構成している。より具体的に、このフィル
20 タ部は、所定の周波数帯域を通過させるバンドパスフィルタを FIR フィルタや IIR フィルタ等のデジタル回路によって構成したものであって、本実施形態では、100 Hz ~ 1500 Hz の周波数帯域の信号成分を通過させるように設定している。

25 打撃音パターン情報格納部 104 は、ボールとラケッ

トとの当たり具合等による音の変化を、スマッシュした時の打撃音やフォアハンドストロークをした時の打撃音などストロークの種類に分類してパターン化した打撃音パターン情報として所定の周波数とその周波数における振幅値とを対応づけて格納するものであって、前記外部記憶装置 12 や内部メモリ 13 の所定領域に形成している。なお、ボールがコートで跳ね返るときの音など、ボールとラケットとが当たることにより発生する音以外の音をパターン化して格納していてもよい。

10 打撃時刻情報特定部 105 は、打撃音パターン情報格納部 104 に格納される打撃音パターン情報と前記音響情報取得部 103 で取得した音響情報とに基づき打撃時刻を特定するものである。

より具体的には、この打撃時刻情報特定部 105 は、
15 前記音響情報取得部 103 で取得した音響情報を 2048 点（≒ 0.046 秒）単位で且つ 128 点（≒ 0.0029 秒）間隔で開始時刻を順次ずらしながら FFT 処理を行い、各時刻における周波数領域に変換した音響情報の周波数特性パターンを前記打撃音パターン情報格納部 104 に格納される打撃音パターン情報と照合するよう
20 うに設定している。そして、これらの照合の結果、音響情報の周波数特性パターンと前記打撃音パターン情報とが一致していれば、その一致した時刻をこのボールの打撃時刻 t_a と特定し、特定した打撃時刻 t_a を画像内容認識部 106 に出力するように動作する。なお、本実施形
25

態では、音響情報の周波数特性パターンと打撃音パターン情報との一致を相関関数を用い、この相関関数が予め設定した閾値より大きな値を示す際に一致したとみなすように設定している。

- 5 画像内容認識部 106 は、前記ドメイン要素抽出部 101 で抽出したコートライン及びネットラインと選手位置情報とボールの位置と、前記打撃時刻情報特定部 105 で特定した打撃時刻 t_a における使用用具の位置と、前記ルール情報格納部 102 に格納されるルール情報と
- 10 に基づき、その映像情報が示す選手の動作を含む画像内容を認識するものである。

より具体的には、第 10 図に示すように、ボール位置の検出が成功した最後の、または、次に続く N 点から適当な軌道を推定することにより特定された打撃時刻 t_a

15 におけるボール位置 $P_i(t_a)$ を求め、このボール位置 $P_i(t_a)$ と選手位置とから、例えば、このボール打撃時刻 t_a においてボールが選手を囲む外接四角形上部の識別ラインより上方にあれば “overhead_swing”、選手重心に対してフォアサイドあるいはバック

20 サイドにあれば、それぞれ “forehand_swing”、“backhand_swing” と選手の動作を判定するように設定している。なお、前記識別ラインは、選手外接四角形の縦方向の長さに応じ一定の比率で決まる選手領域上部に設定している。

- 25 次に、本実施形態の画像認識装置の動作を第 11 図に

示すフロー図を用いて説明する。

- まず、選手のプレイ中の動作が映った映像情報からコートラインとネットラインとをそれぞれ抽出し（ステップ S 1 0 1）、映像情報からこれらコートライン及びネットラインを除去した 2 値画像を用いて選手位置情報を抽出する（ステップ S 1 0 2）。そして抽出した選手位置情報に基づき映像情報からボールを抽出する（ステップ S 1 0 3）。一方、前記コンテンツからボールの打撃時に発生する打撃音を含む音響情報をフィルタ部でフィルタリングして取得し（ステップ S 1 0 4）、フィルタリングして取得した音響情報に対し、所定の間隔で開始時刻を順次ずらしながら F F T 処理を行う（ステップ S 1 0 5）。そして、各時刻における F F T 処理により周波数領域に変換して得た打撃音候補データの周波数特性パターンを打撃音パターン情報格納部に格納される打撃音パターンと照合し（ステップ S 1 0 6）、照合した結果、打撃音候補データの周波数特性パターンと前記打撃音パターン情報とが一致していれば（ステップ S 1 0 7）、その一致した時刻をこのボールの打撃時刻 t_a と特定し（ステップ S 1 0 8）、一致していなければ（ステップ S 1 0 7）、次の時刻における打撃音候補データの周波数特性パターンと前記打撃音パターンとの照合を行う（ステップ S 1 0 6）。そして、特定した打撃時刻におけるボール位置と選手位置及びルール情報に基づき、例えば、第 1 0 図に示すように、ボールが選手に重なったり隠蔽されたり

する場合等特に画像認識に不具合を生じるような場合でも、フォアハンドスイング動作を表す“forehand_swing”、バックハンドスイング動作を表す“backhand_swing”、オーバーヘッドスイング動作を表す“overhead_swing”の3種類の動作で認識することができる（ステップS109）。

以上のようにして、映像中において使用用具が選手やネット等の障害物に重なったり隠蔽されたりして使用用具の位置特定が困難なとき等、映像情報だけでは画像認識が困難な場合でも、音響情報取得部が取得した打撃音を含む音響情報に基づき打撃時刻情報特定部がその打撃音の発生した打撃時刻を特定し、さらに、特定した打撃時刻と選手のプレイ中の動作が映った映像情報と当該ルールを行うためのルール情報とに基づき画像内容認識部が確実に選手の動作を特定するため、例えば、重なりや隠蔽によるフォアハンドスイング、バックハンドスイング及びオーバーヘッドスイングの識別誤りといった映像情報だけでは避けられなかった認識誤りを回避することができる画像認識に優れた画像認識装置を比較的安価に提供することができる。ボールと選手とが重なったり隠蔽されたりしない場合での画像認識が好適に行えることは言うまでもない。

また、打撃音以外のノイズ音が取得した音響情報に含まれていた場合でも、フィルタ部でこのノイズ音を好適にフィルタリングできるため、ロバスト的で認識率の高

い画像認識が可能となる。

さらに、音響情報から複数個の打撃音候補データを取得し、これに基づき打撃時刻を特定するように構成しているため、正確な打撃時刻の特定が可能となる。さらに
5 このとき、複数個の打撃音候補データを、前後の打撃音候補データ間で相互に重なる時刻を有するようにしているため、誤って打撃時刻を特定できないといった不具合も防止できる。

なお、本実施形態において、コンテンツをテニス番組
10 とし、その映像情報から抽出するドメイン要素を使用施設情報をコートライン及びネットラインとしたが、コンテンツが他のスポーツ番組等に替われば抽出する使用施設情報もこれらから変更されることは言うまでも無い。また、選手位置情報、使用用具情報に関してもこれと同
15 様に変更することとなる。

さらに、テレビジョン受像機TVやVTR等の記録再生装置を用いて表示する放映中の番組や記録媒体などに記録されるスポーツに関するコンテンツから、特徴的な試合中の選手の動作を認識するように構成していたが、
20 例えば、スタジアムで当該スポーツを撮影した放映される前の素材映像やインターネット上にアーカイブ化されている映像情報から、特徴的な試合中の選手の動作を認識するなど、画像認識を行うコンテンツの対象となる媒体は本実施形態に限られるものではない。

25 また、画像内容認識部106において映像情報が示す

選手の動作を含む画像内容を、フォアハンドスイング動作を表す“forehand_swing”、バックハンドスイング動作を表す“backhand_swing”、オーバーヘッドスイング動作を表す“overhead_swing”の3種類の動作で認識するように構成していたが、ボール位置や選手位置との関係等に基づき、選手がその場に留まる動作を表す“stay”、選手の移動動作を表す“move”を認識することもできる。また、ルール情報格納部102に格納するルール情報を、
5 選手のさまざまな動作を含めたより複雑なものに定義して格納すれば、画像内容認識部106においてより複雑な選手の動作を認識することも可能である。

なお、 $b_x \times b_y$ のボールを含む所定のテンプレート $T_b(x, y)$ を用いて映像情報からボールを抽出するように構成していたが、このテンプレートを用いずにボールを抽出しても構わない。
15

また、音響情報取得部103にバンドパスフィルタで構成されるフィルタ部を設けたが、バンドパスフィルタ以外のフィルタを用いる実施態様も考えられる。さらに、
20 通過させる周波数帯域も100Hz～1500Hzに限られるものではない。

なお、音響情報取得部103が、ボールの打撃時に発生する打撃音等を含む音響情報を前記コンテンツから分解能16bit、サンプリングレート44.1kHzにてサンプリングして取得するように設定していたが、
25

解能及びサンプリングレートの設定はこれに限られるものではない。

また、前記音響情報取得部 103 で取得した音響情報を、前記打撃時刻情報特定部 105 が 2048 点（ ≈ 0.046 秒）単位で且つ 128 点（ ≈ 0.0029 秒）間隔で開始時刻を順次ずらしながら FFT 処理を行うように設定していたが、FFT 処理を行うポイント数等はこれに限らず他の値に設定しても構わない。

さらに、音響情報の周波数特性パターンと打撃音パターン情報との一致を相関関数を用い、この相関関数が予め設定した閾値より大きな値を示す際に一致したとみなすように設定していたが、音響情報の周波数特性パターンと打撃音パターン情報との一致をみなす方法は他にも考えられる。

その他、各部の具体的構成についても上記実施形態に限られるものではなく、本発明の趣旨を逸脱しない範囲で種々変形が可能である。

産業上の利用可能性

上述したように、本発明によれば、映像中において使用用具が選手やネット等の障害物に重なったり隠蔽されたりして使用用具の位置特定が困難なとき等、映像情報だけでは画像認識が困難な場合でも、音響情報取得部が取得した打撃音を含む音響情報に基づき打撃時刻情報特定部がその打撃音の発生した打撃時刻を特定し、さらに

- 、特定した打撃時刻と選手のプレイ中の動作が映った映像情報と当該ルールを行うためのルール情報とに基づき画像内容認識部が確実に選手の動作を特定するため、例えば、重なりや隠蔽によるフォアハンドスイング、バック
- 5 クハンドスイング及びオーバーヘッドスイングの識別誤りといった映像情報だけでは避けられなかった認識誤りを回避することができる画像認識に優れた画像認識装置を比較的安価に提供することができる。

請求の範囲

1. ネット等の障害物で区画された領域間で対戦するスポーツにおける選手の動作を、そのスポーツの放映中の番組又は放映前の素材映像やVTR等の記録媒体に記録されるコンテンツから認識する画像認識装置であって、前記コンテンツから少なくとも一方の選手のプレイ中の動作が映った映像情報を取得する映像情報取得部と、前記領域間を移動し当該スポーツの得点のカウント対象となるボール等の使用用具の打撃時に発生する打撃音等前記映像情報と同期した音響情報を前記コンテンツから取得する音響情報取得部と、この音響情報取得部が取得した音響情報に基づき前記使用用具を打撃した打撃時刻を特定する打撃時刻情報特定部と、当該スポーツを行うためルール情報を格納するルール情報格納部と、前記映像情報取得部で取得した映像情報と前記打撃時刻情報特定部で特定した打撃時刻における使用用具の位置と前記ルール情報格納部に格納されるルール情報とに基づき、その映像情報が示す選手の動作を含む画像内容を認識する画像内容認識部とを備えていることを特徴とする画像認識装置。

2. 前記打撃時刻情報特定部が、前記音響情報が、所定のレベルより大きな値を示した際に、この大きな値を示す時刻を打撃時刻と特定することを特徴とする請求の範

図第 1 項記載の画像認識装置。

3. 前記音響情報取得部が、所定の周波数帯域を通過させるフィルタ部を備え、前記音響情報がこのフィルタ部を通過したものであることを特徴とする請求の範囲第 1

5 項又は第 2 項記載の画像認識装置。

4. 前記フィルタ部が、バンドパスフィルタで構成されることを特徴とする請求の範囲第 3 項記載の画像認識装置。

5. 前記音響情報から抽出した前記打撃音を含む所定時間
10 間を有する打撃音候補データに基づき、前記打撃時刻情報特定部が打撃時刻を特定することを特徴とする請求の範囲第 1 項、第 2 項、第 3 項又は第 4 項記載の画像認識装置。

6. 一の時刻の打撃音候補データとその次の時刻の打撃
15 音候補データとが相互に重なる時刻を有するように前記音響情報から複数個の打撃音候補データを抽出し、この複数個の打撃音候補データに基づき、前記打撃時刻情報特定部が打撃時刻を特定することを特徴とする請求の範囲第 1 項、第 2 項、第 3 項又は第 4 項記載の画像認識装
20 置。

7. 前記複数個の打撃音候補データを各々同一のデータ
長を有するように構成すると共に、複数個の打撃音候補データを前記音響情報から一定時間間隔で抽出するように構成していることを特徴とする請求の範囲第 6 項記載
25 の画像認識装置。

8. 前記使用用具と選手がプレイ中常に持って使用するラケット等の用具との当たり具合等による音の変化をパターン化した打撃音パターン情報を格納する打撃音パターン情報格納部を備え、前記打撃時刻情報特定部が、この打撃音パターン情報格納部に格納される打撃音パターン情報と前記音響情報とに基づき前記打撃時刻を特定することを特徴とする請求の範囲第1項、第2項、第3項、第4項、第5項、第6項又は第7項記載の画像認識装置。
9. 前記映像情報取得部が、ネット等の障害物や前記領域及びその領域外の境界を示す境界線等の使用施設情報と選手の位置を示す選手位置情報と前記領域間を移動し当該スポーツの得点のカウント対象となる使用用具情報とを映像情報から抽出するドメイン要素抽出部を備えていることを特徴とする請求の範囲第1項、第2項、第3項、第4項、第5項、第6項、第7項又は第8項記載の画像認識装置。
10. 前記選手位置情報が、選手とその選手がプレイ中常に持って使用する用具とを含む領域を示す位置情報であることを特徴とする請求の範囲第9項記載の画像認識装置。
11. 前記ドメイン要素抽出部が、該ドメイン要素抽出部で抽出した使用施設情報に基づき前記映像情報から選手位置情報を抽出することを特徴とする請求の範囲第9項又は第10項記載の画像認識装置。

1 2 . 前記ドメイン要素抽出部が、該ドメイン要素抽出
部で抽出した使用施設情報及び選手位置情報に基づき前
記映像情報から使用用具情報を抽出することを特徴とす
る請求の範囲第9項、第10項、第11項記載の画像認
5 識装置。

1 3 . 前記使用施設情報、前記選手位置情報、前記使用
用具情報及び前記ルール情報が、画像抽出の対象となる
スポーツ種目に関する知識に基づくものであることを特
徴とする請求の範囲第9項、第10項、第11項又は第
10 1 2 項記載の画像認識装置。

1 4 . コンピュータを作動させて、ネット等の障害物で
区画された領域間で対戦するスポーツにおける選手の動
作を、そのスポーツの放映中の番組又は放映前の素材映
像やVTR等の記録媒体に記録されるコンテンツから認
15 識する画像認識装置を稼動するプログラムであって、当
該コンピュータを、前記コンテンツから少なくとも一方
の選手のプレイ中の動作が映った映像情報を取得する映
像情報取得手段と、前記領域間を移動し当該スポーツの
得点のカウント対象となるボール等の使用用具の打撃時
20 に発生する打撃音等前記映像情報と同期した音響情報を
前記コンテンツから取得する音響情報取得手段と、この
音響情報取得手段が取得した音響情報に基づき前記使用
用具を打撃した打撃時刻を特定する打撃時刻情報特定手
段と、当該スポーツを行うためルール情報を格納するル
25 ール情報格納手段と、前記映像情報取得手段で取得した

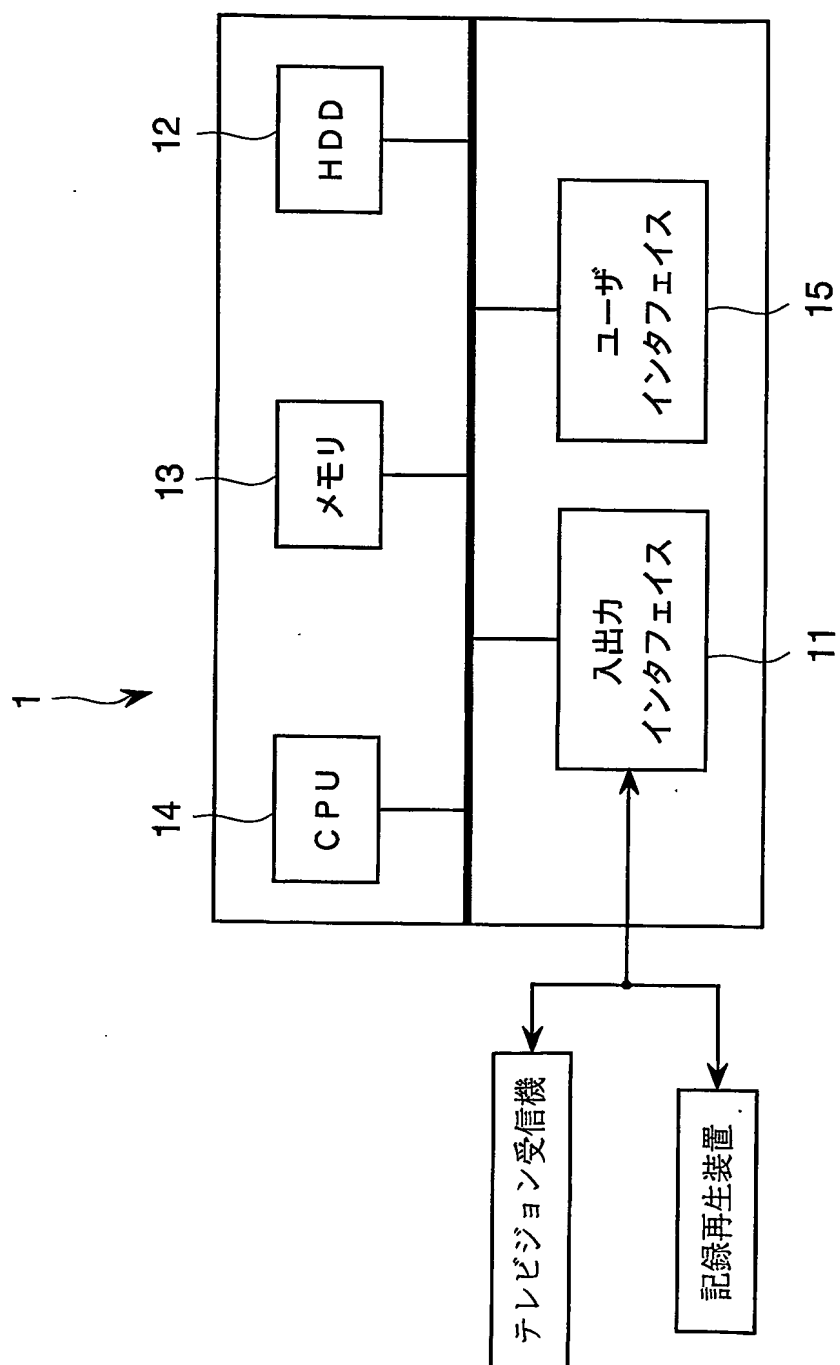
映像情報と前記打撃時刻情報特定手段で特定した打撃時刻における使用用具の位置と前記ルール情報格納手段に格納されるルール情報とに基づき、その映像情報が示す選手の動作を含む画像内容を認識する画像内容認識手段

5 として機能させることを特徴とする画像認識プログラム

。

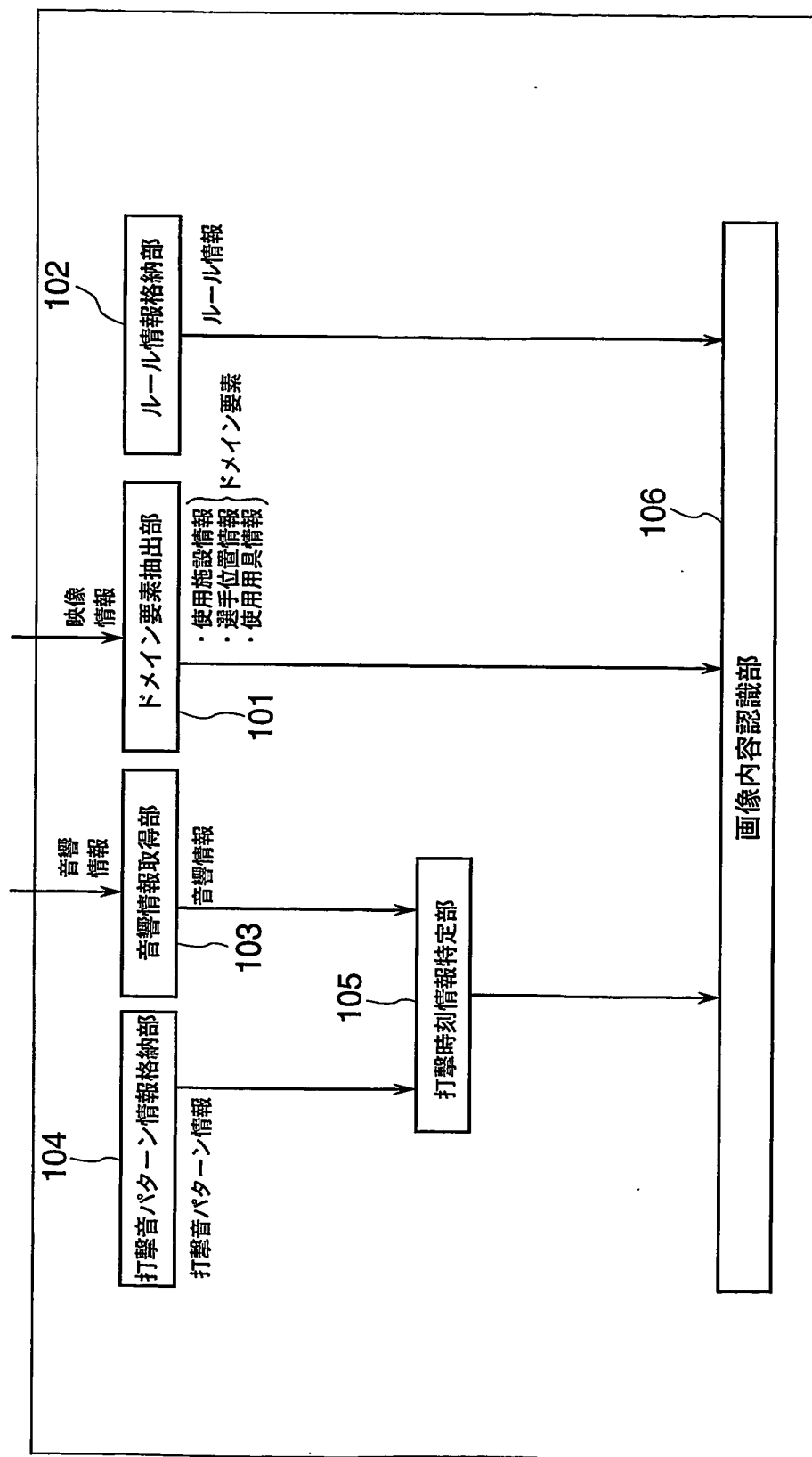
1/11

第1図



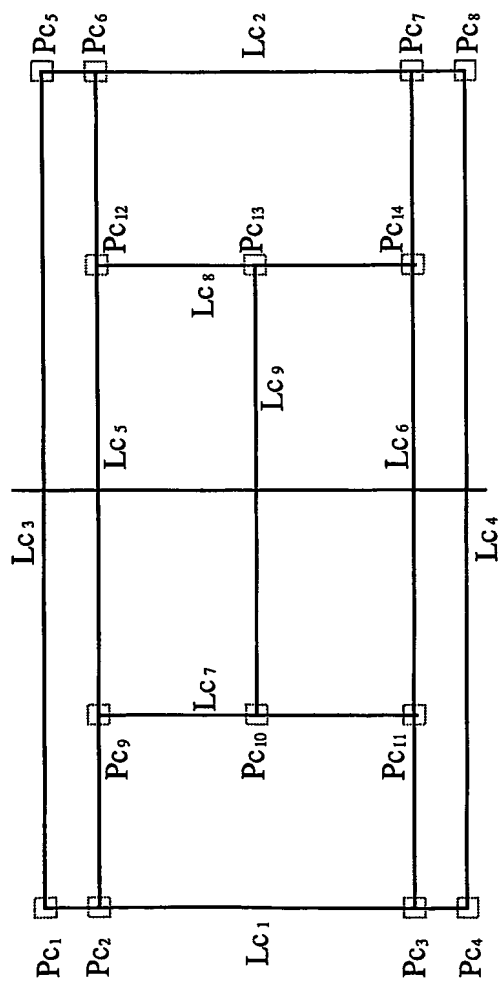
2/11

第2図



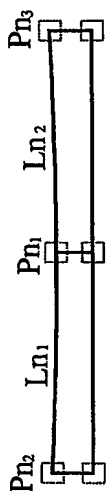
3/11

第3図



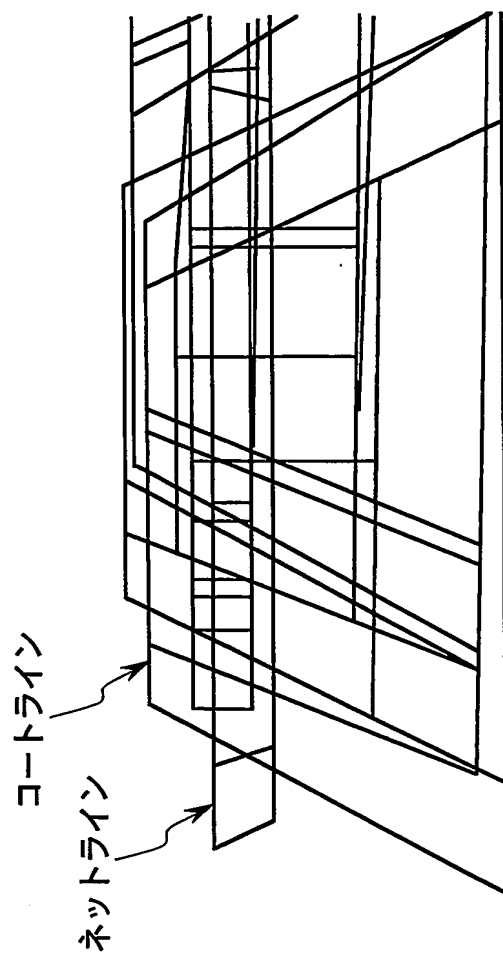
4/11

第4図

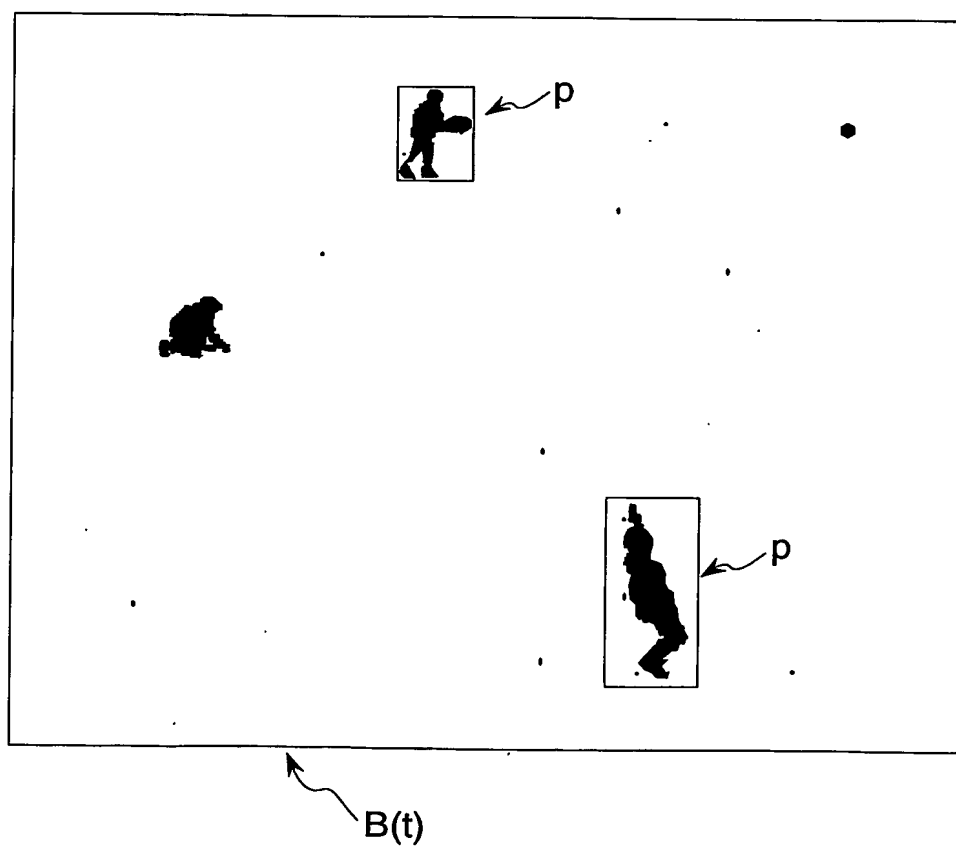


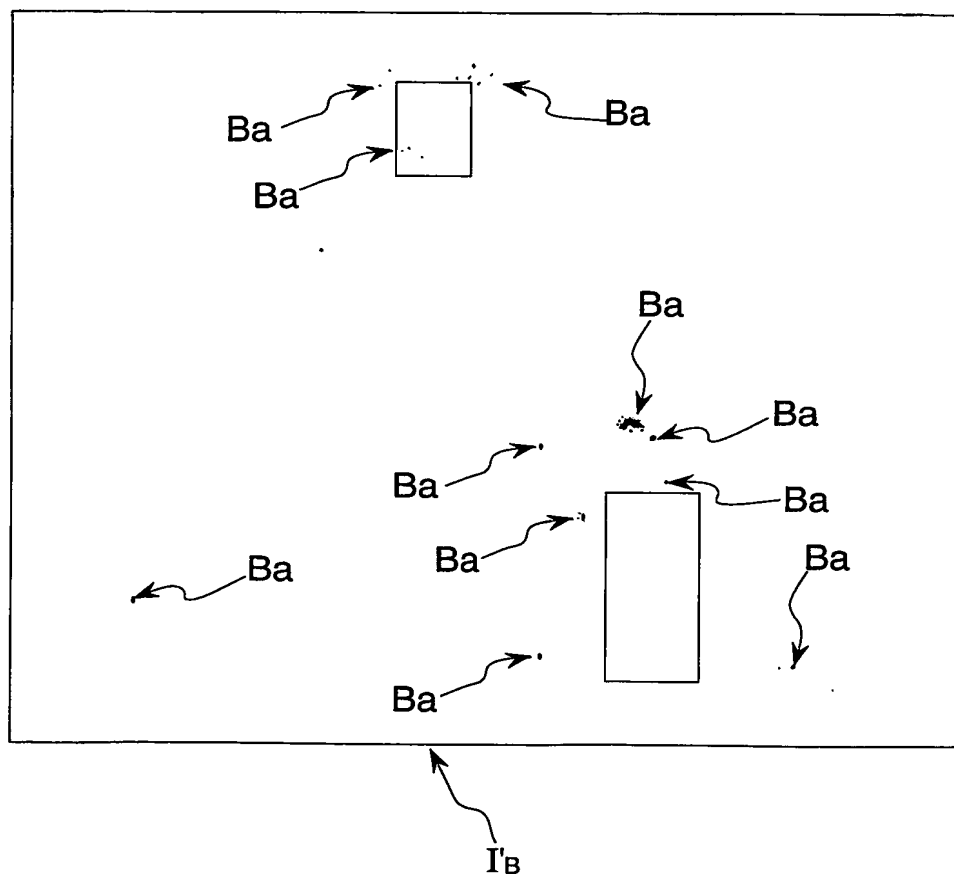
5/11

第5図

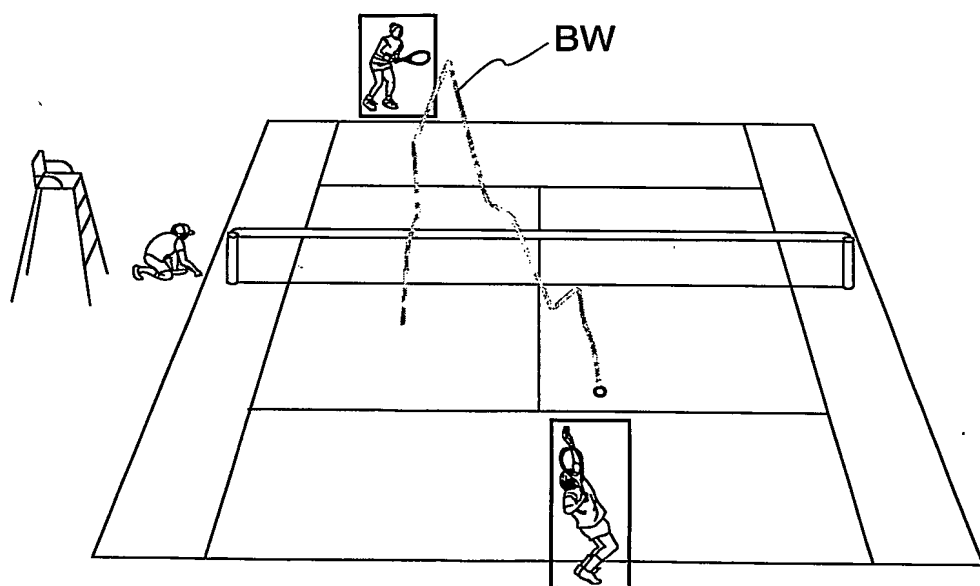


6/11
第6図



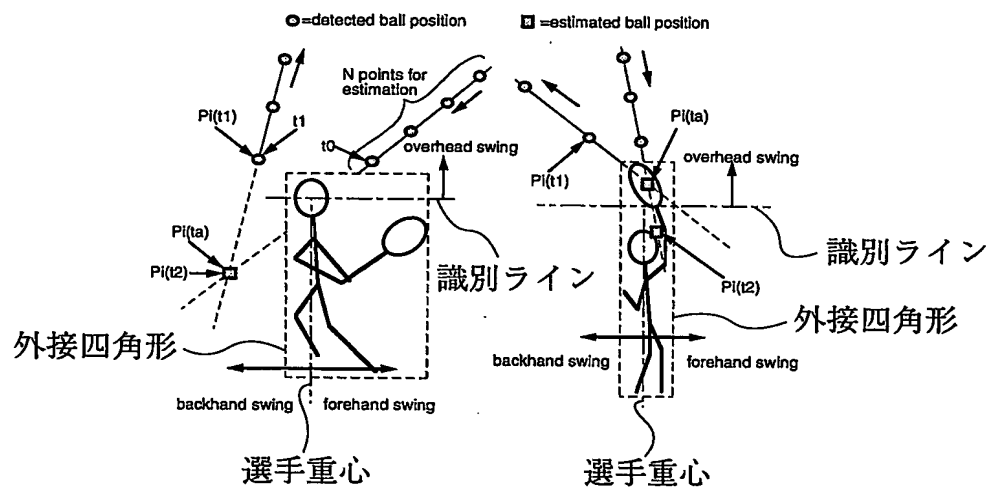
7/11
第7図

8/11
第8図

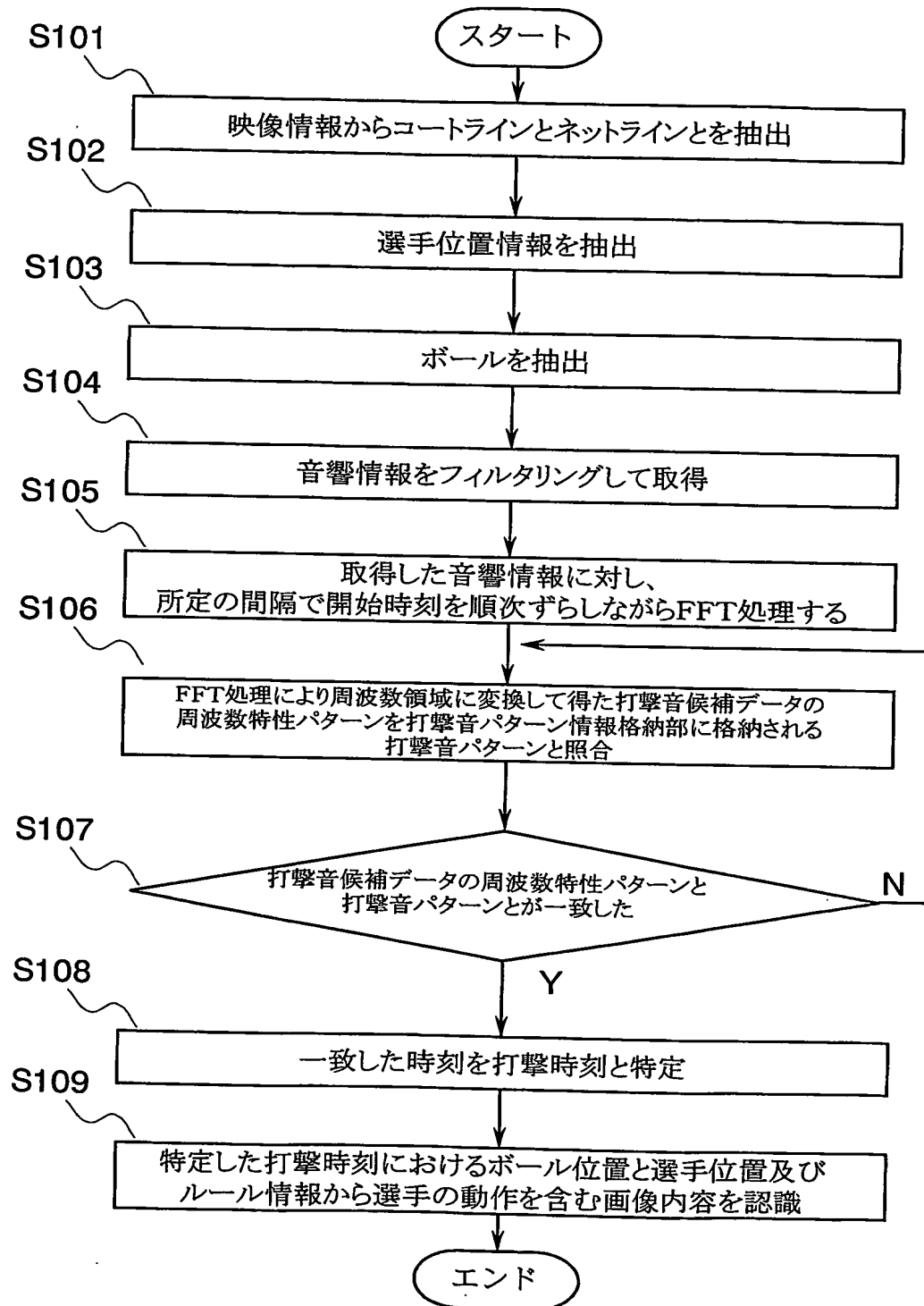


9/11
第 9 図

ルール情報インデックス	ルール情報
サービス	サービスを開始する直前、サーバーはベースラインのネットに向かって後方、センターマークとサイドラインの仮想延長線間に両足とも地面につけて立つ。手でボールを空中いずれの方向にでもほうり投げ、そのボールが地面に落ちるまえにラケットで打つ。サービスはラケットとボールが接した瞬間に完了したものとみなす。
コートライン上の落球	コートライン上に落ちたボールは、そのコートラインによって区切られたコート内に落ちたとみなされる。
・ ・ ・	・ ・ ・

10/11
第10図

バックハンドストローク動作 スマッシュ動作

11/11
第 1 1 図

INTERNATIONAL SEARCH REPORT

International application No.

T/JP02/07826

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ G06T7/20, G06T7/60, G06T1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ G06T7/20, G06T7/60, G06T1/00, H04N5/91-5/956, H04N7/18,
G06F17/30, A63B69/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1926-1996	Toroku Jitsuyo Shinan Koho	1994-2002
Kokai Jitsuyo Shinan Koho	1971-2002	Jitsuyo Shinan Toroku Koho	1996-2002

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Hisashi MIYAMORI, "Eizo-Onkyo Joho no Kyocho niyoru Naiyo Kensaku no tame no Dosa Shikibetsu Seido no Kaizen", Information Processing Society of Japan. Kenkyu Hokoku, 08 March, 2002 (08.03.02), Vol.2002, No.26, pages 89 to 94	1-14
Y	JP 11-339009 A (Sony Corp.), 10 December, 1999 (10.12.99), Par. No. [0028] (Family: none)	1-14
A	EP 820788 A2 (Kabushiki Kaisha ASOBOU'S), 28 January, 1998 (28.01.98), & US 6071002 A & JP 9-313660 A	1-14

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
28 August, 2002 (28.08.02)Date of mailing of the international search report
10 September, 2002 (10.09.02)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G06T 7/20, G06T 7/60, G06T 1/00

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G06T 7/20, G06T 7/60, G06T 1/00, H04N 5/91-5/956, H04N 7/18, G06F 17/30, A63B 69/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926-1996年
 日本国公開実用新案公報 1971-2002年
 日本国登録実用新案公報 1994-2002年
 日本国実用新案登録公報 1996-2002年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	宮森恒, 映像・音響情報の協調による内容検索のための動作識別精度の改善, 情報処理学会研究報告, 2002. 03. 08, Vol. 2002, No. 26, pp. 89-94	1-14
Y	JP 11-339009 A (ソニー株式会社) 1999. 12. 10, 段落0028 (ファミリーなし)	1-14
A	EP 820788 A2 (K.K. ASOBOU'S) 1998. 01. 28 & US 6071002 A & JP 9-313660 A	1-14

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」 同一パテントファミリー文献

国際調査を完了した日

28. 08. 02

国際調査報告の発送日

10.09.02

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

松浦 功

5H

9181

電話番号 03-3581-1101 内線 3531